Putting GPT-3's Creativity to the (Alternative Uses) Test

Year: 2022 | Citations: 110 | Authors: C. Stevenson, I. Smal, M. Baas

---

Abstract

AI large language models have (co-)produced amazing written works from newspaper articles to novels and poetry. These works meet the standards of the standard definition of creativity: being original and useful, and sometimes even the additional element of surprise. But can a large language model designed to predict the next text fragment provide creative, out-of-the-box, responses that still solve the problem at hand? We put Open AI's generative natural language model, GPT-3, to the test. Can it provide creative solutions to one of the most commonly used tests in creativity research? We assessed GPT-3's creativity on Guilford's Alternative Uses Test and compared its performance to previously collected human responses on expert ratings of originality, usefulness and surprise of responses, flexibility of each set of ideas as well as an automated method to measure creativity based on the semantic distance between a response and the AUT object in question. Our results show that -- on the whole -- humans currently outperform GPT-3 when it comes to creative output. But, we believe it is only a matter of time before GPT-3 catches up on this particular task. We discuss what this work reveals about human and AI creativity, creativity testing and our definition of creativity.